

A bibliometric methodology for identifying interdisciplinary and collaborative publications

Heather K. Moberly¹, Jessica R. Page², Gregory K. Youngen³, Barbara J. Hamel⁴

¹*Texas A&M University, College Station, Texas, USA*

²*Ohio State University, Columbus, Ohio, USA*

³*Indiana State University, Terre Haute, Indiana, USA*

⁴*University of Wisconsin, Madison, Wisconsin, USA*

Corresponding author: Moberly, hmoberly@library.tamu.edu

Abstract

Introduction

This presentation describes a bibliometric methodology to define an ancillary journal list as a complement to a core journal list. Although it can be applied to any discipline with a core journal literature, this case study data set is research published by faculty at the 28 American Veterinary Medical Association accredited veterinary schools in the United States. This ancillary list identifies interdisciplinary and collaborative publications by analyzing the non-core subject literature.

Methods

Eleven years of citation data were collected from *Web of Knowledge* and exported to Excel. Data in several fields were normalized, pivot tables were created, and data were uploaded into Many Eyes visualization tool. The result sets were compared to the current core veterinary serials list. The images from both the pivot tables and Many Eyes showed clear trends in the data for each school and across schools.

Results and Conclusions

Overall, 56 percent of articles were published in the core veterinary journals. Bradford's Law and a Bradford-Zipf plot show an enormous breadth of veterinary publications.

Key words: Bibliometrics, Library Collection Development, Interdisciplinary Communication, Data Mining, Data Visualization, Veterinary Medicine

Introduction

The multi-disciplinary nature of animal health and science research requires access to a broad array of medical and biological literature. While a newly updated "core" list of journals covering veterinary medicine has been published, the ancillary literature frequently used in broader veterinary medical sciences and animal health research is more elusive. (1) Specialty journals in fields like dentistry, ophthalmology, cardiology, etc., are often as important to specialists in animal health as they are in human medicine. The varied species of the patients makes the identification of the ancillary veterinary medicine literature more complex than in human medicine. This paper accompanies the 2014 EAHIL presentation. Additional information, including additional graphics, are available open access at and <http://crl.acrl.org/content/early/2013/06/28/crl13-476.full.pdf+html> and <http://bit.ly/RAmUOd>.

The need for precise identification of the ancillary literature is critical, especially for institutions where large medical research libraries or extensive medical collections are absent. This would include research-intensive programs in nursing, pharmacy, dentistry, and a number of other medical fields possibly taught outside formal colleges of medicine.

Ensuring access to the most relevant and up-to-date peer-reviewed journal literature is an essential function of the library but identification of ancillary literature supporting institutional research is an elusive target. User needs vary by institution, by the ever-changing internal research priorities, and over time. An ongoing process of assessment is essential when reviewing journal subscriptions to ensure the collection's relevancy to its users.

One must also factor in the publishing trends in the specialty fields, the local research strengths and needs, and the ever-present budgetary restrictions. While several key journals in medicine (*JAMA*, *The New England Journal of Medicine*, et al.) have application to all fields of medicine, including veterinary, researchers' individual specializations call for a systematic assessment of use and need.

This bibliometric study is designed to map and identify the published works of veterinary researchers and to identify collaborations among veterinary schools. It builds on a pilot study that was initiated to inform collection decisions at veterinary libraries. (2)

Bibliometrics

Bibliometrics is the "application of mathematical and statistical methods to books and other means of communication." (3) Bibliometric studies of published communications within a field are used to quantitatively or qualitatively describe the research occurring within a discipline and assess research productivity within a field. (4) Types of bibliometric studies include: Descriptive analyses of article characteristics within a field, e.g., the change over time in the publication rate or average number of authors per paper within a field; comparisons of author productivity; journal productivity studies including journal rank within a field; identification of scientific collaborations; and citation analyses.

Veterinary medicine

The American Veterinary Medical Association's Council on Education accredits veterinary medical education programs in the United States. The data for this study, presented here as a case study to illustrate the methodology, are 11 years of publication citations from veterinary school faculty and staff at the 28 veterinary schools in the United States accredited at the time of the study. Veterinary medicine, its research publications, and its library resource needs are interdisciplinary. Veterinary school facilities, libraries, studies, and personnel who write papers vary. This variety complicates capturing publication data because of the non-standard way authors might represent their affiliation in published papers.

Authors from veterinary schools can include faculty, adjunct faculty, staff, post-doctoral fellows, residents, interns, and students. Study topics can include clinical, applied, or basic science and the studies might focus on a single discipline or be interdisciplinary. Studies might be conducted solely within the veterinary school, with collaborators from other colleges at the same institution,

or with other institutions. Facilities can include veterinary school departmental laboratories and clinics, private practice veterinary clinics, corporate or government research facilities, veterinary or human medical teaching hospitals, research centers, agricultural or veterinary extension, or agricultural experiment stations.

Methods

Database selection

The veterinary medical literature is indexed in many abstracting and indexing databases. (5-7) *Web of Knowledge* was selected for this study for several reasons. (8) It was a product available to all this paper's co-authors, allowing them to divide the labor and work more efficiently. *Web of Knowledge* has broad, multidisciplinary coverage, including a large number of academic subject areas beyond the health or life sciences, which helped in identifying both publications in a larger number of journals and more interdisciplinary collaborations. The bibliographic records compiled in *Web of Knowledge* were relatively clean and consistent; while there were some irregularities, the extracted data required less clean-up than data from some other sources. (9) Lastly, institutional affiliation data was available for all authors, rather than just for the first author, allowing identification of collaborations among authors from different institutions.

In 2011, at the time searches were conducted for this study, Thomson-Reuters referred to its search platform as *Web of Knowledge*, and to a particular set of databases available on that platform as *Web of Science*. In January 2014, the platform was renamed *Web of Science* and the databases were renamed *Web of Science Core Collection*. For consistency throughout this paper the authors will adhere to the original name *Web of Knowledge*.

Search term selection

Each author developed and tested searches for seven of the twenty-eight veterinary schools using the same settings within the *Web of Knowledge*. The indexes searched were the Science Citation Index (SCI), Social Sciences Citation Index (SSCI), and Arts & Humanities Citation Index (A&HCI). The searches were limited to articles published in the 11 year time period from 2000 to 2010. The "Lemmatization" feature was active during the searches.

Trial search strings were developed for each veterinary school based on the institution's zip codes and the truncated term "vet*". These initial, simple searches were often incomplete or contained errors. Some errors were due to institutions having more than one zip code such as one assigned to the main campus and another covering a campus building at another location. Others errors were due to inconsistencies and inaccuracies in the author-submitted affiliation data *Web of Knowledge* used to index articles. This included not only erroneous zip codes, but variations in how authors' referred to their departments, colleges, teaching hospitals, research institutes, or state diagnostic laboratories. Inaccurate results were also found due to *Web of Knowledge*'s use of the abbreviation, "vet", for both "veterinary" and "veteran".

The SAME operator rather than the AND operator was used in searches to connect zip code terms with institutional terms. In *Web of Knowledge* address searches, the SAME operator is used to find records where the terms separated by the operator appear in the same address. This prevented the search from returning articles where one author was affiliated with the university

in the search and another author had “vet” in her affiliation, but no authors were actually at the target veterinary school.

After developing and running these initial searches, feedback from colleagues at the libraries serving each veterinary school was sought. This led to search string enhancements to capture more of the work conducted at each veterinary school regardless of the author-provided institutional description. Consulting with local veterinary medical librarians allowed this study’s authors to ensure that each location, program, institute and department within a veterinary school was included in the search. In some cases, local librarians confirmed the accuracy of initial searches. In others, they suggested changes, which added to the final search strings’ complexity and accuracy.

Data Analysis and Visualization

The revised searches yielded a final data set of 51,721 records. This data set represented the publication output from the 28 AVMA accredited veterinary schools during 2000 to 2010 as recorded in the *Web of Knowledge*. It should be noted that the *Web of Knowledge* does not include all veterinary journals. (2) The “Analyze Results” feature was used to create separate subsets for source titles, subject areas, institutions, countries/territories, and publication years. This generated lists; for example, all journals (source titles) where researchers at a particular veterinary school published at least two articles. Each subset, except publication years, was sorted by the number of records, largest to smallest, within the subset. The publication years subset was sorted chronologically, newest to oldest.

Once the searches were complete, 168 tab-delineated files were exported. Five tab-delineated files were exported for each of the 28 veterinary schools: source title, subject areas, institution, countries/territories, and publication year. In addition, the complete citation information for all articles published by each school was exported. All of these analysed subset files for a specific veterinary school were combined resulting in one Microsoft Excel workbook file for each veterinary school. Each Excel file had six worksheets: year, subject, source, institutions, countries, and articles. The result was 28 files, one for each school.

Those 28 files were combined into a master Excel file with a worksheet for each exported parameter: journals, subjects, collaborators, countries, year, and articles. The data were normalized as needed because, although *Web of Knowledge* is a relatively clean database, there was variation among the records. In particular, journal titles in the source names subset, institution names in the institution subset, and country names in the countries subset were standardized. The study’s authors normalized the data as needed.

After the data were standardized, a list of all journal titles were extracted from the articles worksheet and copied to a new worksheet. The journal titles on this list also occurring on the core list of veterinary journals were marked, and all others were marked as non-core. (1) Unfortunately, not all fields have an established core list. Although it hasn’t been updated recently, the Collection Development Section of the Medical Library Association provides a subject-based resource list identifying core resources in selected health sciences disciplines. (10)

Excel was used to create pivot table reports for each parameter (journals, subjects, institutions, etc.). Pivot tables allow users to summarize and rearrange long lists of data from spreadsheets, and generate totals without using calculations. For example, after creating a pivot table of the journal data, this paper's authors were able to view at a glance the number of publications from each veterinary school published in each journal, and to rank journals by publications at each veterinary school or overall. They were also able to filter this data by whether the journals were on the core list of veterinary serials or not. Using pivot tables in this way allowed them to view and identify trends in the data.

Selected data was uploaded to IBM's visualization tool Many Eyes. (11) Visualization tools help people understand the significance of data by placing it in a visual context. Many Eyes was chosen because it is freely available, is easy to use, offers the ability to upload a dataset, and enables visualization of data in different ways. Many Eyes' cloud-based storage and display allows easy data and visualization sharing. Many Eyes consolidates a number of visualization tools (i.e., the Wordle word cloud generator, network diagramming, simple pie charts, and histogram chart types) into a single platform. This made it easy to produce many visualization types that illustrate the linkages in the co-authorship and collaboration networks in the data extracted from Web of Knowledge. (see visualizations associated with this project at <http://bit.ly/RAmUOd>).

Many Eyes does not allow extensive customization. The already normalized dataset from *Web of Knowledge* needed additional modification, albeit minimal. Country data posed the first challenge. Although the data were normalized, Many Eyes has a list of countries built in and could not interpret all countries exported from *Web of Knowledge*. Records exported from *Web of Knowledge* coded for England, Northern Ireland, Scotland, and Wales were combined and coded as the United Kingdom. Records coded for St. Kitts and Nevis were added to the records for the West Indies Associated States. Records for Yugoslavia and Serbia were combined and coded as Serbia.

Many Eyes offers a variety of visualization types. Each visualization type has specific requirements for the data type and format. Many Eyes provides useful help files and descriptions for each visualization type that makes these requirements clear. Once a visualization type was selected, appropriate columns of data were chosen and when necessary rows and columns were transposed.

In addition to viewing and visualizing overall trends in the data, Bradford analysis was performed to identify a core set of journals where veterinary school researchers publish the most and sets of auxiliary journals. (4) The distribution of publications within journals in the data set was quantified using Bradford's Law of Scattering. Bradford's Law states that, for a certain number of Bradford zones, where each zone has about the same number of articles, the number of journals in each consecutive zone is an exponential expression of the number of journals in the preceding zone (i.e., the ratio is $1:k:k^2:k^3\dots$). For a set of articles, the number of articles in each Bradford zone is the same, but the number of journals will differ. Typically, a set of articles fitting Bradford's Law has a core zone with few journals and larger zones with exponentially more journals.

The Bradford analysis is useful for dividing a ranked list of journals into discrete sets indicating journals that are essential, recommended, supplemental and peripheral. This could be a useful tool for collection management and provides an opportunity for local analysis.

Results and Discussion

Analysis of the research output of a defined veterinary population highlighted trends in publishing and institutional collaboration and singled out unique qualities of individual institutional programs. The methodology has application to fields beyond veterinary medicine.

Trends in Publishing

Bradford's law is valid for this data set, which comprises four Bradford zones (see Figure 1). The first zone identifies the journals within which the veterinary researchers most frequently publish i.e. the “essential” journals. Journals in the second zone could be considered “recommended” and the third zone “supplemental” since the frequency of publication is progressively smaller. The fourth zone in this study is large, as is typical of datasets where articles are broadly published across a large number of journals. Journals in this fourth zone are considered “peripheral.” Together the four zones indicate an enormous breadth of veterinary publications.

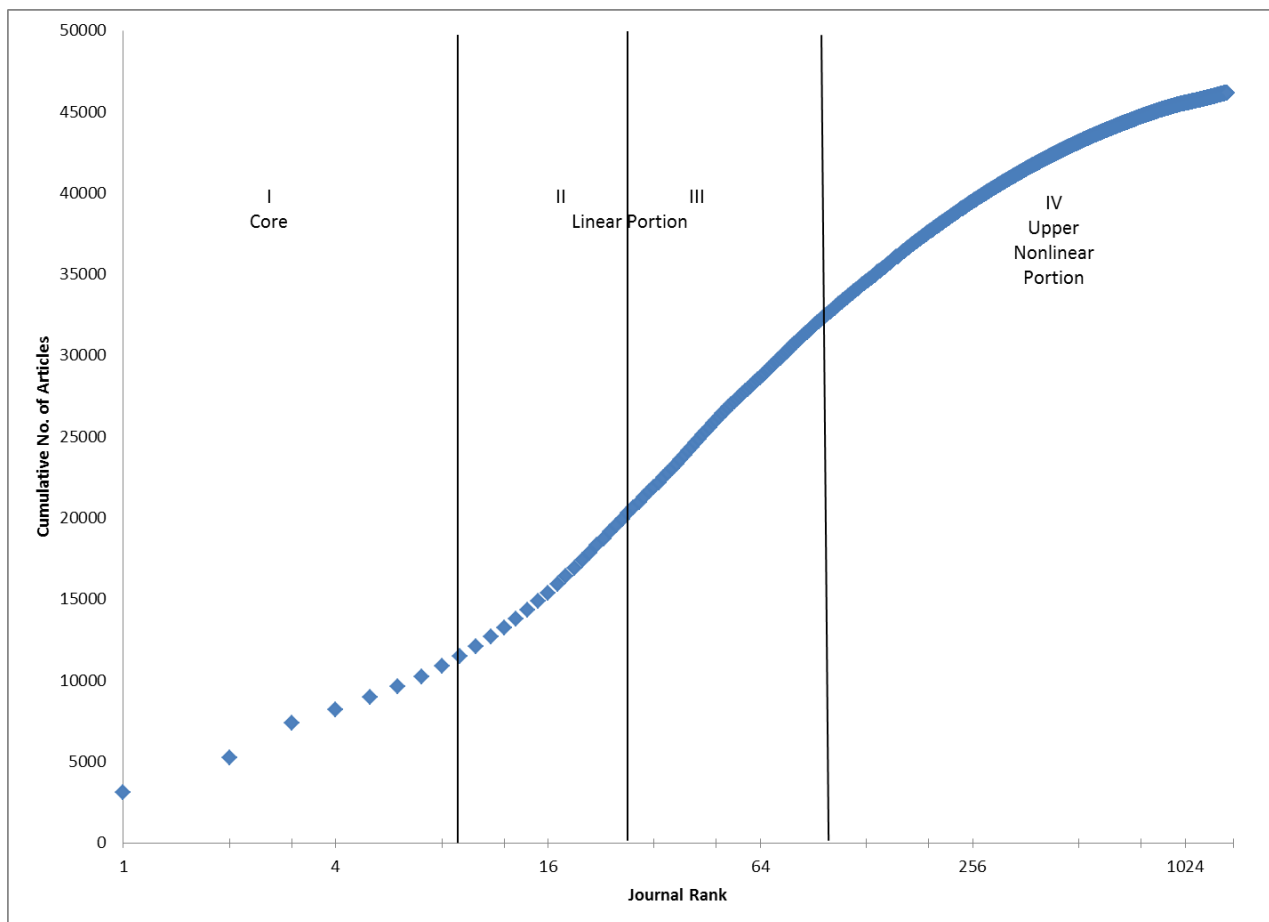


Figure 1. Bradford-Zipf plot of the cumulative number of articles against journal rank. The S-shaped curve is consistent with the Bradford model.

Bradford analysis in this study was limited to journals where researchers from at least one veterinary school published at least two articles during the study period. This culled a large data set encompassing 1,349 journals and 46,172 articles. If the complete set of 51,721 articles had been used, the fourth zone would have been even larger, with a total of 2,700 journals, broadening to an even larger periphery of journals.

Table 1. Division into Bradford zones of journals in which authors from at least one veterinary school published at least two articles.

Zones	No. of journals	No. of articles	Cumulative no.	Cumulative %	Description
1	9	11466	11466	25%	Producing between 614 and 3086 articles
2	27	11566	23032	50%	Producing between 279 and 609 articles
3	93	11565	34597	75%	Producing between 60 and 276 articles
4	1220	11575	46172	100%	Producing from 2 to 59 articles
Total	1349	46172			

By comparing the research output of a defined set of authors in veterinary colleges against the core journal literature of their discipline, the non-core or ancillary journal literature of the discipline was identified. In veterinary medicine, 56 percent of the research output was in articles published in the core veterinary journals. The other 44 percent were published in non-veterinary journals covering basic sciences and medical specialties demonstrating the interdisciplinary nature of the field. The same methodology applied to research coming out of schools of pharmacy might have shown a range of ancillary journals in chemistry, biochemistry, and toxicology.

Knowing which journals researchers in a particular program or at a given institution publish in can help with collection development decisions. Resources of relatively greater or lesser importance can be identified and decisions made based on outcomes. Every program is different and ancillary journal lists can be customized for each. There are also implications for funding. Some ancillary journals may be better aligned with the scope of another campus library's collection or another librarians collection responsibilities.

Trends in Collaboration

Analysis of the veterinary literature revealed trends in intra-institutional, inter-institutional, and international collaboration. Institutional pairs were identified based on number and percentage of publications shared in common. Veterinary researchers frequently collaborated with researchers in other departments within the same institution. Researchers at one veterinary college frequently collaborated with researchers at other veterinary colleges but they also collaborated with institutions lacking veterinary programs such as the United States Department of Agriculture. Most collaborations were with institutions within the United States but more than 1,000 articles had co-authors from the United Kingdom, Canada, and Germany.

Many Eyes is effective in creating visualizations showing the complex collaborations with campus research departments, sometimes revealing unknown partnerships and often validating known partnerships within institutions, with other institutions, and internationally.

Identifying Unique Qualities of Institutional Programs

Analysis of publication output reveals focus and strength of an individual institution's program or that of a peer institution and provides a means of comparison. Distinct profiles for each program emerge and can show whether research output is growing or shrinking over time and how it is being funded.

Word cloud visualizations in Many Eyes are particularly useful for representing an institution's subject concentration. Comparison of word clouds for different institutions shows clear programmatic distinctions. In the case of veterinary schools, Mississippi State University's word cloud shows catfish as the most prominent research subject where Iowa State University in contrast focuses on viruses and pigs.

Deeper knowledge of the programs and disciplines served by the library enhances communication with faculty and researchers in those areas and help strengthen the library-department relationship. Sharing results of departmental research output with program administrators both demonstrates the value of the library and justifies departmental resource and funding needs.

Conclusion

This veterinary medicine case study relied a relatively small and well-defined base of researchers but the methodology is applicable to larger or more loosely defined disciplines. Once the citation data is obtained from the Web of Knowledge through carefully constructed searches, Bradford analysis provides a means of identifying the essential, recommended, supplemental, and peripheral journals in any field of research, customized for local strengths, needs, and budgets. Many Eyes is a useful tool for representing complex collaboration and subject data visually.

References

1. Ugaz AG, Boyd CT, Croft VF, Carrigan EE, Anderson KM. Basic list of veterinary medical serials, third edition: using a decision matrix to update the core list of veterinary journals. *Journal of the Medical Library Association*. 2010;98(4):282-92.
2. Page JR, Moberly HK, Youngen GK, Hamel BJ. Exploring the Veterinary Literature: A Bibliometric Methodology for Identifying Interdisciplinary and Collaborative Publications. *College & Research Libraries*. Accepted: June 1, 2013. Anticipated Publication Date: September 1, 2014.
3. Pritchard A. Statistical bibliography or bibliometrics? *Journal of Documentation*. 1969;25(4):348-9.
4. Andrés A. *Measuring Academic Research: How to Undertake a Bibliometric Study*. Oxford: Chandos Publishing; 2009.

5. Grindlay DJC, Brennan ML, Dean RS. Searching the Veterinary Literature: A Comparison of the Coverage of Veterinary Journals by Nine Bibliographic Databases. *J Vet Med Educ.* 2012;39(4):404-12.
6. Jacso P. As we may search – comparison of major features of the Web of Science, Scopus, and Google Scholar citation based and citation-enhanced databases. *Curr Sci.* 2005;89(9):1537-47.
7. Neuhaus C, Daniel H-D. Data sources for performing citation analysis: an overview. *Journal of Documentation.* 2008;64(2):193-210.
8. Web of Science [Internet]. Thomson Reuters. 2014 [cited 15 Apr 2014]. Available from: <http://apps.webofknowledge.com/>.
9. Youngen G, Gullen A. The Diverse and Interdisciplinary Nature of Veterinary Research: An Analysis of Research Trends at Six Midwestern Colleges of Veterinary Medicine. United States Agricultural Information Network (USAIN) 12th Biennial Conference; May; Purdue University, West Lafayette, IN2010.
10. Medical Library Association Collection Development Section. Subject-Based Resource List 2006 [updated 2009 Jul 13; cited 2014 Apr 15]. Available from: <http://colldev.mlanet.org/resources/subjectlist.htm>.
11. IBM Research. Many Eyes 2007 [cited 2014 Apr 15]. Available from: <http://www-958.ibm.com/software/data/cognos/manyeyes/>.